



INSTITUTO TECNOLÓGICO DE ESTUDIOS SUPERIORES DE MONTERREY
ESCUELA DE INGENIERÍA Y CIENCIAS
CONCENTRACIÓN INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA DE DATOS
MONTERREY, NUEVO LEÓN

TC3006C.103. INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA
DE DATOS I

**RETO ANÁLISIS DEL CONTEXTO Y LA NORMATIVIDAD: INFORME
SOBRE EL USO DEL CONJUNTO DE DATOS SEER BREAST CANCER**

PORTAFOLIO ANÁLISIS

MARIO JAVIER SORIANO AGUILERA A01384282

6 DE SEPTIEMBRE DE 2023

Índice

1. Introducción	2
2. Investigación de la normativa asociada al conjunto de datos	2
3. Explicación del medio utilizado para los datos y el cumplimiento normativo	2
3.1. Explicación del medio utilizado para los datos y el cumplimiento normativo	2
3.2. Cumplimiento normativo y ético de la herramienta	3
3.3. Escenarios de falta ética	3
3.4. Inclusión en un repositorio	3
4. Conclusiones	4
5. Referencias	4

1. Introducción

Este informe tiene como objetivo proporcionar una evaluación exhaustiva del conjunto de datos SEER Breast Cancer, que contiene información sobre pacientes con cáncer de mama en los Estados Unidos. El conjunto de datos fue obtenido de la actualización de noviembre de 2017 del Programa SEER del Instituto Nacional del Cáncer (NCI), el cual ofrece estadísticas de cáncer basadas en la población. Este conjunto de datos específico incluye a pacientes diagnosticadas con carcinoma ductal infiltrante y carcinoma lobulillar infiltrante de mama durante el período de 2006 a 2010. Se excluyeron pacientes con tamaño de tumor desconocido, ganglios linfáticos regionales examinados desconocidos, ganglios linfáticos regionales positivos desconocidos y pacientes cuya supervivencia fue inferior a 1 mes. Esto resultó en la inclusión de un total de 4024 pacientes.

Los datos incluyen detalles como la edad en el momento del diagnóstico, raza, estado civil, estadio del tumor y los ganglios linfáticos, grado de diferenciación del tumor, tamaño del tumor, estado de estrógeno y progesterona, número de ganglios linfáticos examinados y positivos, meses de supervivencia y estado vital. Estos datos pueden ser utilizados para investigaciones y análisis relacionados con el cáncer de mama, pero es importante abordarlos con cuidado para evitar discriminación o uso indebido de la información.

2. Investigación de la normativa asociada al conjunto de datos

En los Estados Unidos, la privacidad y la seguridad de los datos de salud están reguladas principalmente por la Ley de Portabilidad y Responsabilidad del Seguro de Salud (HIPAA) y la Ley HITECH (Ley de Tecnología de la Información para la Salud Económica y Clínica). HIPAA (Ley de Portabilidad y Responsabilidad del Seguro de Salud) HIPAA, promulgada en 1996, establece reglas y estándares para proteger la privacidad y la seguridad de la información de salud de los pacientes. Estas reglas incluyen la Regla de Privacidad de HIPAA, que rige el uso y la divulgación de la información de salud protegida (PHI); la Regla de Seguridad de HIPAA, que establece estándares de seguridad para la información de salud electrónica (ePHI); y la Regla de Notificación de Brechas de HIPAA, que requiere notificar a individuos y autoridades en caso de violaciones de seguridad significativas que afecten a la PHI. Ley HITECH (Ley de Tecnología de la Información para la Salud Económica y Clínica) La Ley HITECH, promulgada en 2009, refuerza las disposiciones de privacidad y seguridad de HIPAA. Introduce sanciones más severas por el incumplimiento de HIPAA y exige notificaciones adicionales en caso de violaciones de seguridad significativas.

3. Explicación del medio utilizado para los datos y el cumplimiento normativo

3.1. Explicación del medio utilizado para los datos y el cumplimiento normativo

El conjunto de datos SEER Breast Cancer se obtuvo del Programa SEER del NCI y se cargó en U-BRITE para el "AI against CANCER DATA SCIENCE HACKATHON". La fuente original de los datos es "SEER Breast Cancer Data.^{en} IEEE Dataport, y se menciona que los datos se proporcionan bajo la licencia "Attribution 4.0 International (CC BY 4.0)". Para cumplir con la normativa y la licencia, es fundamental

asegurarse de que se incluya la atribución adecuada en cualquier uso o redistribución de los datos. Además, se deben implementar medidas de seguridad y privacidad, como la anonimización de datos y el control de acceso, para proteger la información de salud de los pacientes.

3.2. Cumplimiento normativo y ético de la herramienta

Cumplimiento normativo y ético de la herramienta Para garantizar el cumplimiento normativo y ético al utilizar el conjunto de datos SEER Breast Cancer, es necesario abordar varias consideraciones clave: Anonimización adecuada de los datos: Para proteger la privacidad de los pacientes, se debe llevar a cabo una anonimización efectiva, eliminando cualquier información de identificación personal y asegurándose de que los datos no puedan vincularse a individuos específicos. Seguridad de datos: Dado que los datos de salud son sensibles, es esencial implementar medidas de seguridad de datos, como el cifrado y la gestión de acceso, para prevenir el acceso no autorizado. Evitar sesgos en el análisis: Se debe prestar especial atención para evitar cualquier sesgo en el análisis de datos y al interpretar los resultados. Esto garantiza la equidad y la imparcialidad en la investigación. Consentimiento Informado: Si se planea utilizar datos de pacientes en investigaciones futuras, es importante obtener el consentimiento informado de los pacientes o cumplir con las regulaciones de consentimiento aplicables. Uso Responsable: Utilizar los datos para mejorar la atención médica, la investigación médica y la toma de decisiones clínicas en beneficio de los pacientes y la salud pública. El dataset cumple con los puntos anteriores, aunque se está limitado a cómo es que esta información publica es usada.

3.3. Escenarios de falta ética

Existen escenarios potenciales de falta ética al utilizar datos de salud, como el conjunto SEER Breast Cancer. Algunos de estos escenarios incluyen: Discriminación: El mal uso de los datos para discriminar a ciertos grupos de pacientes con base en su historial médico o características personales. Decisiones perjudiciales: Tomar decisiones perjudiciales para la salud de las pacientes basadas en el análisis de datos, en lugar de utilizar los resultados para mejorar la atención médica. Para prevenir estos escenarios de falta ética, es crucial establecer políticas y protocolos claros para el uso responsable de los datos y promover su utilización en beneficio de la salud de los pacientes.

3.4. Inclusión en un repositorio

Una vez completados los informes y análisis basados en los datos del conjunto SEER Breast Cancer, es fundamental agregarlos a un repositorio académico o plataforma de investigación. Asegúrese de proporcionar la atribución adecuada a las fuentes y la licencia, y siga los requisitos de la licencia CC BY 4.0 al compartir estos datos con otros. Además, documente claramente cómo se cumplen las regulaciones y los estándares éticos en sus análisis y aplicaciones basadas en estos datos.

4. Conclusiones

El conjunto de datos SEER Breast Cancer es una valiosa fuente de información para la investigación médica y científica, pero su uso conlleva responsabilidades normativas y éticas. Al comprender y cumplir con las regulaciones, garantizar la privacidad y la seguridad de los datos y promover un uso ético de estos recursos, podemos aprovechar al máximo su potencial para avanzar en la comprensión y el tratamiento del cáncer de mama mientras protegemos los derechos y la privacidad de los pacientes.

5. Referencias

Rights, O. F. C. (2023, 28 agosto). HIPAA home. HHS.gov.

<https://www.hhs.gov/hipaa/index.html>

ONC | Office of the National Coordinator for Health Information Technology. (2023, 1 julio).

<https://www.healthit.gov/>

Text - H.R.1 - 111th Congress (2009-2010): American Recovery and Reinvestment Act of 2009. (2009, febrero 17).

<https://www.congress.gov/bill/111th-congress/house-bill/1/text>

Breast cancer. (2022, 8 agosto). Kaggle.

<https://www.kaggle.com/datasets/reihanenamdari/breast-cancer>

JING TENG. (2019). SEER Breast Cancer Data. IEEE Dataport.

<https://dx.doi.org/10.21227/a9qy-ph35>

Creative Commons — Attribution 4.0 International — CC BY 4.0. (s. f.).

<https://creativecommons.org/licenses/by/4.0/>